RECOMMENDATION SYSTEM IN SELECTING COURSE OF PUBLIC UNIVERSITY IN MALAYSIA USING K-NEAREST NEIGHBOUR

THOI WEN BIN

BACHELOR OF COMPUTER SCIENCE

UNIVERSITI MALAYSIA PAHANG

Universiti Malaysia PAHANG
Engineering • Technology • Creativity

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Bachelor in Computer Science (Computer Systems & Networking) with Honours.

(Supervisor's Signature)

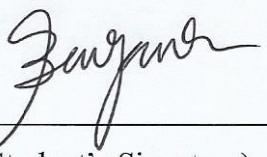
Full Name : DR. JUNAIDA BT SULAIMAN

Position : SENIOR LECTURER
FACULTY OF COMPUTER SYSTEMS
& SOFTWARE ENGINEERING
UNIVERSITI MALAYSIA PAHANG
[illegible] TUN RAZAK, 26300 GAMBANG, KUANTAN
TEL: [illegible] 2255 FAKS: 09-549 2144

Date : 10/1/2019

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.



(Student's Signature)

Full Name : THOI WEN BIN

ID Number : CA15037

Date : 25 December 2018

RECOMMENDATION SYSTEM IN SELECTING COURSE OF PUBLIC UNIVERSITY IN MALAYSIA USING K-NEAREST NEIGHBOUR

THOI WEN BIN

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Bachelor of Computer Science (Computer Systems & Networking) with Honours

Faculty of Computer Systems & Software Engineering

UNIVERSITI MALAYSIA PAHANG

DECEMBER 2018

ACKNOWLEDGEMENT

Firstly, I would like to express my greatest gratitude to Dr. Junaida Binti Sulaiman, my supervisor who give me support and guidance throughout the research. It is a great honour to do this research under her supervision. During this period, she provided a lot of useful advices to improve my research.

I must express my appreciation to my family and friends as they support me mentally and physically. Their motivation kept me in completing the thesis in time. They also give me a lot of suggestions and ideas to continue my research.

ABSTRAK

Internet mula berkembang dengan kelajuan yang besar pada tahun-tahun kebelakangan ini. Perniagaan telah menemui banyak pelanggan dan pendapatan di internet. Terdapat maklumat dan item yang besar, dan ia menjadi berlebihan. Sistem cadangan dibangunkan untuk mengatasi masalah ini. Kini, pelajar lepasan sekolah khususnya untuk pelajar STPM dan Matrikulasi di Malaysia menghadapi masalah dengan memilih kursus universiti awam yang sesuai. Terdapat 20 universiti awam yang menawarkan 893 kursus sarjana muda di Malaysia. Mereka perlu memohon kursus-kursus yang berminat melalui UPU dalam talian dalam tarikh akhir yang ditetapkan. Di Malaysia, hanya iMASCU yang merupakan sistem cadangan mudah yang menunjukkan dan menyemak kelayakan kursus berdasarkan kelayakan pengguna. Sistem cadangan berasaskan K-Nearest Neighbour akan dilaksanakan untuk menyelesaikan masalah ini. Objektif projek ini adalah mengkaji algoritma dan teknik semasa dalam sistem cadangan untuk memilih kursus; untuk melaksanakan K-Nearest Neighbour dalam sistem cadangan; dan untuk menilai prestasisi sistem cadangan. Dataset projek ini adalah program ijazah yang ditawarkan oleh universiti awam Malaysia. Java akan digunakan untuk melaksanakan projek ini. Data dikumpulkan dari soal selidik yang dibuat kepada pelajar universiti awam. Terdapat enam input dan satu output sasaran dalam latihan dan ujian. Hasilnya menunjukkan bahawa nilai k yang dikehendaki adalah satu dan k-Nearest Neighbor boleh dilaksanakan di pemilihan kursus universiti awam Malaysia.

ABSTRACT

Internet began growing up with immense speed in these recent years. Business field had discovered many customers and income in the internet. There are huge information and items, and it became overload. Recommendation systems is developed to overcome this problem. Nowadays, freshly school leavers especially for STPM and Matriculation college students in Malaysia have trouble with selecting a suitable course of public university. There are 20 public universities offering 893 bachelor courses in Malaysia. They need to apply their interested courses through UPU online within stated deadline. In Malaysia, there is only iMASCU which is the recommendation system that show and check the qualification of courses based on the qualification of user. K-Nearest Neighbour based recommendation system is implemented to solve this problem. The objectives of this project are to study the current algorithm and technique in recommendation systems for selecting courses; to implement k-Nearest Neighbour in the recommendation system; and to evaluate the application of the recommendation system. The data is collected from survey made to current public university students. There are six inputs and one target output in the training and testing sets. The result showed that the desired k value for the data is one and k-Nearest Neighbour can be implemented in the Malaysia public university course selection.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

$\sum$	SUMMATION
$\in$	ELEMENT OF
$\emptyset$	EMPTY SET

LIST OF ABBREVIATIONS

K-NN	K-Nearest Neighbour

CHAPTER 1

INTRODUCTION

1.1 Introduction

Internet began growing up with immense speed in these recent years. Business field have discovered many customers and income in the internet. So that, there are many online shop exists in the internet like Amazon, Lazada and many more. Today, there are huge information and items, and become overload. The seller is very difficult to find out what the user is looking for. This problem can be solved by search engines but only partially solved. However, some search engines not able provide personalization of information. So that, developers found a solution to solve the problem which is recommendation systems (Zhang et al., 2015).

Recommendation system is a software tool or technique that gives recommendations of items for user. Recent year, recommendation systems is being used in many field like movies, music, tourism and others. The purpose of recommendation system is to provide some suggestion to the user who have less experience or ability to assess over many choices. For example, Amazon.com implement recommendation systems to personalize the online store for their customer, so that their customer can receive different suggestion via this recommendation systems. Based on the user's preferences, recommendation system will guess the items that most suitable for user. In education field, there are many recommendation systems appeared in recent years such as Course Selecting Recommendation System (Al-Badarenah & Alsakran, 2016), E-Learning Recommendation System (Tan & Guo, 2008), and Reading Material Recommendation System (Hsu, Hwang, & Chang, 2010).

Nowadays, freshly school leavers especially for STPM and Matriculation college students in Malaysia have trouble with selecting a suitable course of public university. According UPU Online, there are 20 local universities offering 893 bachelor courses in

Malaysia. These fresh school leavers need to review all the courses before apply suitable courses. They will apply through UPU online within stated deadline.

1.2 Problem Statement

Recommendation system had several weaknesses such as the lack of data, changing of user preference, scalability, and privacy (Jain, Grover, Thakur, & Choudhary, 2015). Recommendation system need data to analyse, so that users can only get a good recommendation when the recommendation systems have a huge data set to analyse. On the other hand, user will change their preferences easily because of their purposes or any other factor. So that, recommendation systems will not recommend a suitable item for the user when user changing their preferences.

Course selecting recommendation system had been implemented in other country like Jordan, Korea and others. In Malaysia, there is a recommendation system called iMASCU to show the qualification of courses. iMASCU is a simple recommendation system that show and check the qualification of every courses and apply university course in UPU Online (Shahar, 2018). In other word, iMASCU only show the course that can applied or based on their STPM or Matriculation College result only and not consider their interest and preference. Therefore, the recommendation of iMASCU given will be not suitable for the student that have excellent exam result because he will get many recommendations from the recommendation system and the trouble of reviewing course is not reduce.

This project implement a recommendation system for selecting university course to reduce these problem of STPM students by giving them a suitable suggestion. This project implement k-Nearest Neighbour based recommendation system based on student survey response.

1.3 Objectives

The aim for this project is to develop a recommendation system in selecting course of local university in Malaysia. To achieve the aim of this project, there are some objectives that need to be consider. These objectives are as below,

i) To study the current algorithm and technique in recommendation systems for selecting university courses.

ii) To implement k-Nearest Neighbour in the recommendation system.

iii) To evaluate the application of recommendation system for selecting university course.

1.4 Scope

The scope of this project is divided in two categories, which are:

i) Dataset

The selection of courses is for degree level. The data of degree program is program of Malaysia public university that get from UPU Online. The data of public university also get from UPU Online.

The option used for secondary school and matriculation college subject include Malay Language, English Language, Chinese Language, Science, Mathematics, History, Geography, Biology, Physics, Chemistry, Additional Mathematics, Information Technology, Economics, Business Studies, Accounting and other subjects are categorised as others.

Other than that, choices of hobby include chatting, computer game, cooking, dance, eat, internet, listen to music, play music instrument, reading, singing, sleep, sport, watch drama, movie, and other hobbies are categorised as others.

For the academic discipline of the first choice of university course, the university courses are categorised into two groups, which are Science, Technology, Engineering and Mathematics (STEM), and non-STEM.

There are 73 set of data collected from university students in Malaysia and the 73 set of data is divided into 2 for training testing. 30 set of data used for training process and 43 set of data used for testing process.

REFERENCES

Adeniyi, D. A., Wei, Z., & Yongquan, Y. (2016). Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method. *Applied Computing and Informatics*, *12*(1), 90–108. https://doi.org/10.1016/j.aci.2014.10.001

Al-Badarenah, A., & Alsakran, J. (2016). An Automated Recommender System for Course Selection. *International Journal of Advanced Computer Science and Applications*, *7*(3), 166–175. https://doi.org/10.14569/IJACSA.2016.070323

Baker, F. B. (2001). *The Basics of Item Response Theory. Second Edition.* (Second Edi). Washington: ERIC Clearinghouse on Assessment and Evalution, College Park, MD. Retrieved from https://files.eric.ed.gov/fulltext/ED458219.pdf

Barrington, L., Oda, R., & Lanckriet, G. (2009). Smarter than genius? human evaluation of music recommender systems. *... Symposium on Music ...*, (Ismir), 357–362. https://doi.org/10.1007/978-3-540-72079-9_9

Berry, M. J. A., & Linoff, G. S. (1999). *Mastering Data Mining - The Art and Science of Customer Relationship Management*.

Bhumichitr, K., Channarukul, S., Saejiem, N., Jiamthapthaksin, R., & Nongpong, K. (2017). Recommender Systems for university elective course recommendation. *2017 14th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 1–5. https://doi.org/10.1109/JCSSE.2017.8025933

Bogers, T., & Van Den Bosch, A. (2009). *Collaborative and content-based filtering for item recommendation on social bookmarking websites*. *CEUR Workshop Proceedings* (Vol. 532). https://doi.org/10.1007/978-0-387-85820-3

Facebook, I. (2015). Recommending items to more than a billion people | Engineering Blog | Facebook Code. Retrieved from https://code.facebook.com/posts/861999383875667/recommending-items-to-more-than-a-billion-people/

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. *San Francisco, CA, itd: Morgan Kaufmann*. https://doi.org/10.1016/B978-0-12-381479-1.00001-0

Hsu, C. K., Hwang, G. J., & Chang, C. K. (2010). Development of a reading material recommendation system based on a knowledge engineering approach. *Computers and Education*, *55*(1), 76–83. https://doi.org/10.1016/j.compedu.2009.12.004

Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, *16*(3), 261–273. https://doi.org/10.1016/j.eij.2015.06.005

Jain, S., Grover, A., Thakur, P. S., & Choudhary, S. K. (2015). Trends, problems and solutions of recommender system. *International Conference on Computing, Communication and Automation, ICCCA 2015*, 955–958. https://doi.org/10.1109/CCAA.2015.7148534

Lane, D. (2013). Values of the Pearson Correlation. *Online Statistics Education: An Interactive Multimedia Course of Study*. Retrieved from http://onlinestatbook.com/2/introduction/variables.html

Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, *7*(1), 76–80. https://doi.org/10.1109/MIC.2003.1167344

Park, D. H., Kim, H. K., Choi, I. Y., & Kim, J. K. (2012). A literature review and classification of recommender systems research. *Expert Systems with Applications*, *39*(11), 10059–10072. https://doi.org/10.1016/j.eswa.2012.02.038

Peterson, L. E. (2009). K-nearest neighbor - Scholarpedia. https://doi.org/doi:10.4249/scholarpedia.1883

Poole, D. L., & Mackworth, A. K. (2010). Artificial Intelligence - Foundations of Computational Agents. https://doi.org/10.1016/j.artint.2010.12.004

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-Based Collaborative Filtering Recommendation, (JANUARY). https://doi.org/10.1145/371920.372071

Shahar, F. M. (2018). Business | New Straits Times | Malaysia General Business Sports and Lifestyle News. Retrieved from https://www.nst.com.my/business

Srivastava, T. (2018). Introduction to KNN, K-Nearest Neighbors _ Simplified. Retrieved December 4, 2018, from https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/

Szkuta, B., & Sanabria, L. (1999). Electricity price short-term forecasting using artificial neural networks. *IEEE Transactions on Power*, *14*(3). Retrieved from http://ieeexplore.ieee.org/abstract/document/780895/

Tan, H., & Guo, J. (2008). E-Learning Recommendation System, 430–433. https://doi.org/10.1109/CSSE.2008.305

Wang, Z., Yu, X., Feng, N., & Wang, Z. (2014). An improved collaborative movie recommendation system using computational intelligence. *Journal of Visual Languages and Computing*, *25*(6), 667–675. https://doi.org/10.1016/j.jvlc.2014.09.011

Yoon Ho Choa Soung Hie Kim, J. K. K. (2002). A personalized recommender system based on

web usage mining and decision tree induction. *Expert Systems with Applications*, *23*, 329–342. https://doi.org/10.1016/S0957-4174(02)00052-0

Zhang, L., Bu, Z., Wu, Z., Cao, J., Wu, D., Lu, J., … Biradar, S. (2015). A Personalized E-Learning Based on Recommender System. *Other Conferences*, *1*(2), 417–444. https://doi.org/10.1146/annurev-psych-113011-143823